DOCUMENT RESUME

ED 208 660 FL 012 574

AUTHOR Michopculos, Aristotle
TITLE Language Dominance Testing for Greek Bilingual Children.
PUB DATE May 81
NOTE 54p.; Paper presented at the Annual International Bilingual-Bicultural Education Conference (10th, Boston, MA, May 23-30, 1981). Greek version of appendix has been removed because of poor reproducibility.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Bilingual Education; *Bilingual Students; Diagnostic Tests; Elementary Education; *Greek Americans; *Language Dominance; *Language Tests; Morphology (Languages); Multiple Choice Tests; Student Placement; Syntax; Test Reliability; Test Validity

ABSTRACT

This paper addresses the lack of language dominance assessment instruments and curriculum materials for Greek-speaking children in the U.S. These children need appropriate language screening tests, based on research and data derived from their native language group, for diagnostic and placement purposes. The development of an instrument for language assessment of Greek bilingual children is presented and traced. The subjects for the development of the instrument were 403 male and female first through fourth grade students enrolled in bilingual programs in the U.S. The final instrument consists of 80 items, 40 for the Greek and 40 for the English text, 30 of which are designed to elicit a response to a verbal stimulus using multiple choice pictorial alternatives. Each multiple choice item consists of a word stimulus to be matched to one of a series of five related pictures that follows. The remaining items are verbal stimuli in sentence form, designed to elicit a grammatically and syntactically correct answer to a question. Examination of the instrument in relation to field-test data, internal consistency reliability, test-retest reliability, reliability by school and sex, and validity demonstrated that the instrument possesses high reliability and usable validity values. (Author/JK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED208660

LANGUAGE DOMINANCE TESTING FOR GREEK BILINGUAL CHILDREN

Aristotle Michopoulos

University of Florida

Paper presented at the Tenth Annual International Bilingual/Bicultural Education Conference (Boston, Massachusetts, May 23-30, 1981)

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

X Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. Michopoulos

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL 012 574

INTRODUCTION

As we all know, one of the most significant movements within the American educational system during the past decade is that of bilingual education. This movement was, among other things, the outcome of research (Kobrick, 1972; Rodriguez, 1974), indicating that the treatment of minority school children was characterized by neglect and negation.

One of the publications that helped precipitate this movement was the Coleman Report (Coleman et al., 1965) that portrayed the plight of many minority students by comparing their school achievement, academic self-concept, and dropout figures with those of English-speaking students.

The pedagogical/political effect that this and other similar studies had on bilingual-bicultural education was extremely beneficial (Zirkel, 1974b, p.30), since educators and politicians, for the first time, viewed the cultural and linguistic diversity of minority children as an asset rather than a liability to academic success (Anderson and Boyer, 1970; Saville and Troike, 1971).

This change of attitude, however, would have been of no consequence if the federal government had not provided financial support to bilingual programs, especially after the landmark Supreme Court decision of Lau vs. Nichols (1974). Despite the Lau decision and subsequent development of many bilingual programs, bilingual/bicultural education has not yet achieved its full potential. A serious drawback to its realization is the lack of teaching materials and curriculum designed for bilingual schools.

This lack of materials, in turn, is to a great extent due to the lack of language assessment instruments, a problem which came to light in both the Lau vs. Nichols and the Diana vs. California cases. Many researchers in the field also pointed out that there is a relative paucity of research data on bilingual education in general and on language dominance in particular. Blanco (1977) attributed this paucity to limited federal funding for such research, stating that the United States has spent only one-tenth of one percent of its funding on research to improve the quality of bilingual programs (p.4).

Expressing a similar criticism, Woodford (1977) noted that in addition to other shortcomings of such programs, the "least has been done in the areas of evaluation" (p. 130). He further noted that in most school areas, language dominance determination is based on such "cumbersome and time-consuming processes" (p. 131), as student interviews by teachers or teacher aids.

But even with the use of tests, other problems appear. Harber (1976), for example, cited empirical evidence indicating that "certain psychoeducational placement tests do not measure the learning abilities they were intended to measure when used with linguistically and/or culturally different children" (p. 31).

Discussing the same issue from a different perspective, Eells (1953) also pointed out that for a test to be free of cultural bias it "must be composed of items which deal with materials common to the various subcultures in which it is to be used" (p. 292).

In the past, standardized tests with minority group children resulted in an unfavorable assessment of these children. These discriminatory testing practices, which resulted in multiple traumas

for such children, led to court battles and demands for "conpensatory punitive damages" to plaintiffs in the 1970s. Such cases as Stewart vs. Philips in Massachusetts (1970) or Covarrubias vs. San Diego Unified School District (1971) revealed the damage of discriminatory testing, which placed students in classes for the "educable mentally retarded." These cases, which sought substantial damages ($20,000 per student), "presaged a far-reaching revision of current testing methods" (Oakland, 1977, p. 64). Moreover, they led the justices in other cases, such as the Lau vs. Nichols and Diana vs. California, to call for a moratorium on testing language-minority children with instruments based on white, middle-class American children. Their rationale was that "the reliability, validity, and norming of these tests has been established for the general population rather than for a particular bicultural group" (Bernardoni, 1967, p. 1).

Another result of these judicial proceedings, was an impetus for development of new assessment techniques for minority children. Through funding from DHEW, foundations, or publishing firms, many minority language dominance tests were developed during the 1970s.

Despite the criticism from some writers, on the validity and reliability of a number of such instruments, more progress has been made in this area in the last decade than in the previous 50 years. However, while a multidue of language dominance assessment instruments and curriculum materials have been developed for Spanish-speaking children, a paucity of similar materials is apparent for other minority group languages.

Greek is a case in point. While this minority group is greater in size than other groups, such as Filipinos and Portuguese, which

have such instruments (Lock, Pletcher, and Reynolds, 1978), nothing similar has been developed until recently for this group. According to the U.S. Bureau of Census (1976) over half a million people (542,000) in the United States have Greek as their native tongue. Also, according to the Survey of Income and Education of the U.S. Bureau of Census (1976), approximately 88,000 Greek minority children between the ages of 6 to 18 did not have English as their native tongue or did not use English at home.

These children need appropriate language screening tests for diagnostic and placement purposes. Some teachers of Greek bilingual programs use Spanish language dominance tests, which they have translated into Greek, for such purposes. It is generally agreed, however, that such translations contain numerous potential problems and are usually inappropriate. The best solution to the problem, would, then, be the development of an instrument that would be based on research and data derived from the same language group, Greek in this case, for which it will be developed.

It is true that the academic achievement of the Greek immigrants (Krug, 1976; Moskos, 1980) has been surprisingly high compared to other minority groups. However, some Greek students have been sent to classes for the mentally handicapped, because they scored low on culturally biased achievement tests or because they had to answer a test in a language they did not understand. Such children should be appropriately assessed before they were assigned to classes for the mentally handicapped. Thus, problems, traumas, and stigmas could have been avoided, benefiting both the students and society. The development, then, of such an instrument by this researcher - the

building of which will be discussed in the following pages - aimed at filling this great gap in the area of language assessment of Greek bilingual children and was also a response to federal legislation calling for such instruments.

METHOD

Subjects of the Study

The subjects for the development of the instrument were four hundred and three (403) male and female students attending Grades 1-4 at the Greek Bilingual Program and Greek Parochial School of Tarpon Springs, Florida; the Saint Demetrios Greek Parochial School in New York; and the Budlong Elementary Public School in Chicago.

All the above schools, with the exception of the Tarpon Springs Parochial School, are day schools. The Budlong School is a public school with many bilingual programs, one of which is the Greek Bilingual Program. The student population is basically middle or upper-lower class. The students of the Greek program come mainly from new immigrants of a middle socioeconomic status. Many of these students, therefore, are Greek language dominant. The overwhelming majority of the students of the Soterios Ellenas School in New York, on the other hand, are born in the United States and come from a middle to upper-middle class background. The Saint Demetrios students present a mixture of older and recent parent immigration. Most of these students come from a middle to upper-middle class background.

Criteria used for this selection were locales with a concentration of Greek-speaking people, and with the presence of one or more Greek bilingual schools. The selection of these schools and geographic regions is assumed to be representative of Greek bilingual students of the same age in other parts of the country. This procedure, in turn, of selecting a wide geographical representation, enhanced the

external validity of the instrument, compared to a selection of a more limited scope. Furthermore, the sample size, which exceeded 300 students, was a good representation of the total population and adequate for the development of the instrument.

Development of the instrument began with a large number of items. Subsequent revisions reduced them to 80 items each for the Greek and English versions (60 pictorial and 20 verbal stimuli). The selection of the initial items was based on a random sampling of words appearing in the vocabulary lists of the Grades 1-4 readers for Greek and American children. A pool of more than 60 multiple choice items was derived from such vocabulary lists. These items were divided into four general categories aimed at testing the subjects' knowledge of family, neighborhood, school, and church objects. Likewise, more than 20 Greek and English question items were designed to measure the subjects' command of grammar and syntax in each language.

The final instrument consisted of 80 items, 40 for the Greek and 40 for the English text, 30 of which are designed to elicit a response to a verbal stimulus using multiple choice pictorial alternatives. Specifically, each multiple choice item consists of a word stimulus on the left, followed by a series of five related pictures, one of which is the correct answer to the stimulus. Accordingly, respondents were asked to match the word with the appropriate picture. The remaining items are verbal stimuli in sentence form, designed to elicit a grammatically and syntactically correct answer to a question.

Special care was taken to make the test items reflect the dual history and culture of the subjects. Thus, six items in the multiple choice part of the instrument and two question items differed in the Greek and English components of the tests. The difference reflected the dissimilar sociocultural background of a child raised in Greece as opposed to the U.S.A. For example, item 11 in the multiple choice test depicts a carrot in the English test but an olive in the Greek test, since olives are a much more common item in the Greek diet than carrots; similarly, item 49 depicts the American flag in the English component and the Greek in the Greek.

This general item formulation was presented to the participants of both the general and special Greek colloqium at Florida State University for feedback and recommendations relevant to further improvement. The participants of the general colloqium were approximately 25 doctoral students of the Florida State University bilingual program, many of whom were former bilingual teachers and who represented such diverse language groups as Greek, Spanish, Tagolog, Vietnamese, and Arabic. The participants of the Greek colloqium were eight doctoral students in the Florida State University Greek bilingual program, some of whom were graduates of institutions of higher education in Greece. The instrument was also presented for feedback to other language specialists and experts at Florida State University and elsewhere.

The experts consulted, suggested various changes and/or substitutions in the items, as well as positioning of items within the test. The instrument was revised and again presented to the experts who considered it adequate for its purpose. This examination,

critique, revision, and final approval of the instrument constituted one of the basic steps in establishing its content validity.

After final revision and approval by the panel of reviewers, the pictorial test items were given to an illustrator for illustration. After illustrations were completed, the test was finalized in both Greek and English. The test items were placed in a progressive order of difficulty. It was expected that only the students of higher grades would be able to answer most or all of the items correctly.

Selection of the items was based on the following criteria: (a) the items should take cognizance of the child's ability to use such an item at a particular age, (b) the items should be interesting and enjoyable enough to stimulate the child's response, (c) the items should sample a variety of vocabulary and activities, and (d) the items should include, where possible, the same, similar, or equivalent stimuli in both languages.

Pilot Testing and Results

The pilot testing included 42 subjects from Grades 1-4 of the Greek Bilingual School and the Church School in Tarpon Springs, Florida. Of these 42 test responses - 8 at Grade 1, 10 at Grade 2, 9 at Grade 3, and 15 at Grade 4 - two were found invalid and were excluded from computer analyses.

The tests were subsequently coded and given to the Florida State University Evaluation Service Center for a Standard Item Analysis Program. The main statistics of interest in this analysis were the mean, standard deviation, reliability, and standard error of measurement of each group. These statistics are presented in Table 1.

Table 1

Pilot Test Statistics

		Greek											
		Part A				Part B				Total			
Gr.	n	$\bar{X}$	SD	r_{11}	SEM	$\bar{X}$	SD	r_{11}	SEM	$\bar{X}$	SD	r_{11}	SEM
1	8	20.87	8.76	.88	3.09	1.50	.53	.00	.59	22.38	9.15	.88	3.15
2	9	19.56	16.67	.97	3.04	1.78	1.99	.72	1.06	21.33	18.47	.97	3.20
3	8	28.48	14.27	.95	3.14	1.69	2.85	.83	1.17	30.17	17.32	.96	3.46
4	15	31.53	13.86	.9[illegible]	3.27	1.73	3.37	.88	1.19	33.27	15.56	.95	3.47
		English											
1	8	32.50	6.07	.76	2.99	3.00	2.00	.67	1.14	35.50	7.19	.80	3.20
2	9	41.66	4.20	.66	2.40	8.22	4.17	.83	1.70	49.89	6.66	.81	2.94
3	8	49.63	4.14	.81	1.81	5.00	7.54	.95	1.71	54.63	6.30	.83	2.59
4	15	47.60	13.55	.98	1.76	10.60	7.68	.93	1.98	58.20	16.27	.97	2.77

k = 60 for Part A

k = 20 for Part B

A large number of items in both the Greek and English tests yielded a large difficulty index, i.e. were easy, while a few of them presented a small difficulty index, i.e. were difficult. To remedy this problem the verbal cues of these items were changed. In the item analysis, the number of English test items that did not perform well was larger than the Greek one. The reason for this is that the subjects found many more easy items in the English section of the test than in the Greek one. Thus, almost all changes in the English test were focused on increasing the difficulty of these items.

In Section B of the Greek test, the language component, it was also found that certain items were out of progressive difficulty order and were reordered. The same approach was employed in Section B of the English test.

The individual grade reliabilities of Section A for the Greek test were higher than those of the English, with the exception of Grade 4, where the reliability of the Greek test was .94 and the English .98. The reason for the generally higher reliabilities of the Greek test is attributed to the greater discrimination value of the Greek items, probably caused by their lower difficulty index and greater standard deviation compared to their English counterparts. The reliabilities of Section B, the language component of the test, are generally higher on the English test, but the difference between the two tests here is smaller than in Section A. One notable exception in this section, however, is Grade 1 of the Greek test, where the reliability is zero. This result was due to the fact that most of the first grade students were unable to read or write in Greek, since the Greek alphabet and language are more difficult than English. The effect of the difficulty in reading and writing upon the subjects is also manifested in the low reliability appearing in the same grade and section of the English test. The reason for the higher reliabilities of the English test is that its items were generally easier to the subjects than those of the Greek test. This ease, which is demonstrated in the higher means of the English section, resulted in higher item variance and thus higher reliabilities. Finally, taking into consideration the length of this component -- only 20 items -- one is drawn to the conclusion that its reliability

is quite high.

In comparing the overall reliabilities (Section A and B) of the Greek and English tests, it is apparent that the reliability of the Greek test is higher than that of the English, with the exception of Grade 4, where they are almost equal: .95 and .97.

Examining the means of the test, on the other hand, one easily notes the substantial difference between the Greek and English test means. The difference clearly denotes that the subjects scored higher in English than in Greek, an easily explained fact when one considers that most of the subjects have been in the United States for a number of years and have also been immersed in an English language environment.

Field Testing

Schools and Subjects

After computer analysis of the pilot testing data, and ensuing item revision, the instrument was ready for field testing. The field testing, conducted in early June 1979, included two stages: test and retest. The original number of subjects per grade and school are as follows:

Table 2

Subjects by School and Grade

Grade	School 1	School 2	School 3	Total
1	23	32	39	94
2	25	37	40	102
3	26	11	55	92
4	16	9	48	73
1-4	90	89	182	361

In order to obtain test-retest data, the instrument was administered twice, with an interval of approximately a week between the two test administrations. Of the 361 subjects tested, only 337 responses were used for computer analysis. The remaining 24 tests were excluded from the analyses, either because they were deemed invalid for various reasons, including "Christmas-treeing," insufficient marking, or reading and writing difficulty (10 cases), or because the subjects were not present in both administrations of the instrument (14 cases). The total number of cases excluded from the analyses is quite small compared to the total sample population. The high rate of clean data is partly attributed to favorable cooperation from the school personnel as well as from the subjects.

Data Coding

The response data were coded onto optically scannable answer sheets. In addition, and prior to this, the writer read and marked all Section B test, both Greek and English. At the end, each subject had two opscan sheets of data, one for the test and one for the retest. After the completion of coding, a general check was made on all coded data, test responses and identification data coding, to ensure that there were no missing data and that all marks were properly entered. The rate of missing or incomplete data was less than one percent.

The data were then separated according to grades and an answer key for each test and section was made. The data were given to the Florida State University Evaluation Services to be transferred to a disk file so that the appropriate statistical analyses could be made. These analyses included the following: (a) test reliability

(coefficient alpha), (b) mean, standard deviation, and SEM, (c) test-retest reliability, (d) p values per item, (e) point biserial values per item, and (f) concurrent validity analyses. The analyses were performed for each part and the total of the English and Greek tests, by grade, as well as for the total group of subjects.

RESULTS AND DISCUSSION

The results of the study included the following: (a) analysis of the item pool, (b) item selection for final instrument, (c) characteristics of final instrument, (d) test-retest reliabilities, (e) reliabilities by school and sex and (f) concurrent validity.

Analysis of the Item Pool

The statistical characteristics of the item pool were examined to facilitate selection of items for the final instrument. First, the means and standard deviations of the pupils' scores for each part of the test (Greek and English) were computed. Second, the internal consistency reliability coefficients for the parts were calculated. Third, discrimination indices, correlations between each item and the part score were calculated. The same statistics were calculated for the total 80-item pool. The means, standard deviations, and internal consistency of scores over the items in the pool are shown in Table 3.

Table 3

Characteristics of the Item Pool (N = 302)

		Subtest A			Subtest B			Total Test		
Grade	n	$\bar{X}$	SD	r_{11}	$\bar{X}$	SD	r_{11}	$\bar{X}$	SD	r_{11}
					English Test					
1	85	35.14	8.45	.876	7.40	5.22	.904	42.54	12.76	.927
2	61	38.60	5.62	.765	11.67	4.09	.862	50.27	8.72	.868
3	87	43.87	8.46	.913	14.15	3.83	.846	58.39	10.81	.927
4	69	39.48	4.89	.772	16.97	3.19	.838	56.45	7.04	.853
1-4	302	43.10	8.61	.901	12.57	5.51	.921	55.66	13.20	.941
					Greek Test					
1	85	32.45	13.60	.947	4.93	4.56	.898	37.38	16.67	.955
2	61	42.33	8.81	.890	7.73	4.24	.849	50.07	11.50	.911
3	87	43.03	10.16	.923	10.80	4.20	.821	53.84	12.81	.930
4	69	44.17	10.25	.930	12.26	5.01	.904	56.43	14.36	.951
1-4	302	40.11	12.04	.941	9.04	5.32	.904	49.15	16.05	.954

The data show that the mean performance of subjects was greater on the English items than on the Greek items. Since the content of the items was desinged to be parallel, this difference in performance is probably due to greater competence of the subjects in the English language, which is the primary language in the subjects' schools. Subjects were more variable in their performance on the Greek items than the English items. This finding might be associated with the difference in means and a slight tendency toward ceiling effects.

For example, the largest standard deviation is associated with the smallest mean for total scores on both the English and Greek tests. The items have a substantial degree of internal consistency as reflected in the alpha coefficients.

Item Selection for Final Instrument

In examining the data for internal consistency, item and part analyses showed that the 80-item pool possessed the appropriate statistical characteristics to permit development of a shorter instrument. The final instrument included 40 Greek and 40 English items. Specifically, the selection of the 40 items (30 multiple choice and 10 free-response items) was based on the following criteria: (a) each item should have an acceptable point biserial value (greater than or equal to .30), (b) the parts (English and Greek) of the instrument should be based, item by item, on the same pictorial or question cues, (c) the two parts of the final instrument should possess the closest possible point biserial values, item by item, and items whose Greek or English components might present significantly different point biserial values should be excluded, and (d) the resulting Greek and English parts of the final instrument should present comparable reliabilities. Thus, the final parts of the instrument were designed to be of equal difficulty for a balanced bilingual child.

All items without variance were automatically excluded from consideration in the final selection process. Furthermore, the selection of an item was based on its overall performance and not merely on its performance in selected grades. Based on the foregoing criteria, 40 items were finally selected for both the Greek and English parts of the instrument.

Characteristics of Final Instrument

Reliabilities and other pertinent statistics obtained by the final 40-item instrument are presented in Table 4.

The reliabilities on both the Greek and English parts of the instrument are quite high and only slightly lower than those of the entire item pool (see Table 3). The differences between the reliabilities of the two parts of the instrument are negligible and not statistically significant. The largest difference appears in Grade 4, where the reliabilities were .78 for the English part and .92 for the Greek part. This difference may be attributed to a low variance and high mean for the English part for this grade, which may have caused a ceiling effect, as well as to a restriction in the sample compilation, since the majority of subjects for this grade (47 out of 69) came from the same school. However, the difference between the two reliabilities is not statistically significant.

Table 4

Statistical Characteristics of the Final 40-Item Instrument by Grade Level and Overall

		English Test				Greek Test				Differences between the two parts			
Gr.	n	$\bar{X}$	SD	r_{11}	SEM	$\bar{X}$	SD	r_{11}	SEM	$\bar{X}_d$	SD	r_{11}	SEM
1	85	23.10	8.04	.910	2.41	20.15	9.93	.936	2.38	2.95	8.25	.830	3.38
2	61	30.15	5.70	.843	2.28	28.05	6.62	.865	2.38	2.10	8.13	.840	3.25
3	87	33.98	5.10	.863	1.89	29.27	7.30	.895	2.34	4.70	8.37	.870	3.01
4	69	34.09	4.00	.776	1.92	30.06	8.01	.922	2.24	4.03	7.75	.850	3.02
1-4	302	30.49	7.80	.924	2.18	26.50	9.10	.931	2.37	3.99	8.22	.840	3.29

As shown in Table 4, means for the Greek and English parts are consistently different. This difference, which favors the English part and is larger on the higher grades, is statisfically significant and indicates a greater competence of the subjects in the English language.

The standard deviations of the instrument are generally high compared to those of the 80-item pool (see Table 3), since all items excluded from the final instrument had low or no variance and thus their exclusion had no serious effect on the variability of the instrument. The standard deviations of the Greek part are generally higher than those of the English. These higher standard deviations and accompanying larger variances are the basic reason for the higher reliabilities of the Greek part, since there is a direct relationship between these two statistics. The Grade 4 standard deviation of the Greek part is higher than expected, especially when compared to the standard deviations of Grades 2 and 3. The reason for this high standard deviation may be due to a wide variability in language ability among the subjects of this group.

The standard errors of measurement for both parts are comparably small. These small standard errors of measurement are also reflected in the generally high reliabilities of both parts, since there is an inverse relationship between these two statistics.

The reliabilities of the difference between scores on the English and Greek parts are quite large and stable across grade levels. These values are quite satisfying, because it is the difference between the scores on the English and Greek parts that constitutes the language dominance score. Therefore, these reliabilities

are crucial to the effectiveness of the instrument in reliably differentiating between Greek and English dominant children.

The remaining statistics for the differences, means, standard deviations, and standard errors of measurement are also stable. The standard error of measurement values are higher than those of each individual part because the variances of errors of measurement are added together when the difference between parts is determined.

Test-Retest Reliabilities

The instrument was administered a second time to the subjects, approximately one week after the first administration, to examine the stability of its results over a period of time. Table 5 shows these test-retest reliability results by grade level and overall.

Table 5

Test-Retest Reliabilities

by Grade Level and Overall

			English Part					Greek Part				Differences between the Two Parts			
Gr.	n		$\bar{X}$	SD	r_{11}	SEM		$\bar{X}$	SD	r_{11}	SEM	$\bar{X}$	SD	r_{11}	SEM
1	85	E1	23.10	8.04	.715	2.41	G1	20.15	9.93	.754	2.38	D1 2.95	8.25	.686	3.38
		E2	22.86	8.44		4.47	G2	19.59	9.59		4.80	D2 3.27	7.58		4.24
2	61	E1	30.15	5.70	.863	2.28	G1	28.05	6.62	.821	2.38	D1 2.10	8.13	.860	3.25
		E2	29.75	6.27		2.32	G2	27.58	7.09		2.98	D2 2.17	7.89		2.92
3	87	E1	33.98	5.10	.822	1.89	G1	29.27	7.30	.476	2.34	D1 4.70	8.37	.572	3.01
		E2	33.20	5.34		2.24	G2	28.73	7.39		5.39	D2 4.47	9.06		5.98
4	69	E1	34.09	4.00	.661	1.92	G1	30.06	8.01	.889	2.24	D1 4.03	7.75	.861	3.02
		E2	35.01	4.09		2.37	G2	28.48	9.50		3.14	D2 6.54	8.50		3.15
1-4	302	E1	30.40	7.80	.853	2.18	G1	26.50	9.10	.781	2.37	D1 3.99	8.22	.728	3.29
		E2	29.90	7.96		3.10	G2	25.78	9.37		4.40	D2 4.12	8.38		4.36

The test-retest reliabilities, with the exception of Grade 3 of the Greek part, showed considerable stability of results over time. Additionally, test-retest reliabilities of the difference between the two parts of the instrument are substantial.

Considering the numerous sources of error that can enter into a test-retest administration as compared to a single administration of an instrument, the obtained overall reliabilities of .85 for the English part and .78 for the Greek seem adequately large. The reliabilities of the differences between the English and Greek parts are especially encouraging in view of the difficulties often associated with the use of indices generated from the difference between two scores.

Reliabilities by School

Table 6, which summarizes the reliabilities of the instrument by school, shows that reliabilities were acceptably high both for individual schools as well as overall. Additionally, the reliabilities of the differences between the two parts by individual school and overall are extremely close. The means of the two components are very similar for school 1, but they are higher for the English part in school 3 and especially school 2, denoting that the subjects of schools 3 and 2 had greater competence in English. This is explained by the fact that school 1 is a public school with a large variability of ability among its subjects -- many of which are recent Greek immigrant children -- while schools 3 and 2 are parochial with a more homogeneous, "Americanized" middle class student population. The means of the difference between the two parts express the

numerical difference between the English and Greek parts of the instrument and the higher scoring of the subjects, especially in school 2, on the English part.

Table 6

Reliabilities of the Final Instrument by School(s)

		English				Greek				Differences between the Two Parts			
Sch.	n	$\bar{X}$	SD	r_{11}	SEM	$\bar{X}$	SD	r_{11}	SEM	$\bar{X}$	SD	r_{11}	SEM
1	79	26.31	9.46	.935	2.27	25.99	10.12	.952	2.23	.32	8.45	.850	3.30
2	48	26.89	9.10	.941	2.16	19.30	10.72	.952	2.36	7.59	10.06	.900	3.22
3	175	33.21	5.30	.847	2.07	28.34	7.19	.889	2.37	4.87	6.64	.770	3.19
1-3	302	30.49	7.80	.924	2.18	26.50	9.10	.931	2.37	3.99	8.22	.840	3.29

Finally, the standard deviations of the difference are stable and large across schools. The standard errors of measurement are almost identical among schools, showing that the instrument's performance was similar among schools. This similarity, along with similarities among the standard deviations and reliabilities, show that results of the instrument were similar among different school settings.

Reliabilities by Sex

Sex as a variable did not appear to have an effect on the reliability of the instrument. Table 7 shows the similar reliability values yielded for females and males on both English (.917 vs. .923) and Greek (.925 vs. .942) parts. The table also indicates that the standard error of measurement values for males and females are also small on both parts.

The tests yielded slightly higher standard deviation values for males, especially on the Greek component, resulting in slightly higher but not, statistically, significantly different reliability values for males when compared to females.

Table 7

Reliabilities by Sex

			English				Greek				Difference between the Two Parts			
Sex	Gr.	n	$\bar{X}$	SD	r_{11}	SEM	$\bar{X}$	SD	r_{11}	SEM	$\bar{X}$	SD	r_{11}	SEM
M	1-4	138	29.56	8.27	.923	2.32	24.62	10.03	.942	2.41	4.94	9.53	.880	3.34
F	1-4	164	31.12	7.57	.917	2.12	27.71	8.17	.925	2.12	3.41	6.62	.780	3.11

A difference of means in favor of the female subjects is shown, which is statistically significant on both the English and Greek parts at the $p<.05$ level. The yield of higher scores for female subjects coincides with research findings (Restak, 1979) on sex as a predictor of language acquisition.

In the reliability of difference values between the two parts, a difference, not statistically significant, is shown in favor of the males -- .88 vs. .78. This difference may be attributed to the larger standard deviation of the scores of the males.

Concurrent Validity

Content validity of the instrument was established through a panel of experts. To further investigate the validity of the instrument, however, an examination of its concurrent validity was undertaken. Teacher ratings of the bilingualism of each student were used as an independent criterion measure of language dominance.

Each subject was assigned a code reflecting his/her language classification as designated by the teachers; a code of 1 indicated Greek language dominance, a code of 2 indicated balanced bilingualism, and a code of 3 indicated English language dominance. The teacher ratings were then transferred to the subjects' computer record and a correlation coefficient expressing the teacher/instrument agreement was obtained.

Such correlations are usually low due to the multiple sources of error entering their calculation. Sources of error may arise from a teacher's subjective evaluation of a subject's bilingualism. Such errors may be based on factors such as the length of acquaintance with the subject, teaching experience, teacher's degree of bilingualism, and subject's foreign accent or lack of it, resulting in leniency error, halo effect error, and so on. Such errors increase the likelihood of misclassification of a subject. Another important factor which contributes to the misclassification of a subject is the category of balanced bilinguals, but also bilinguals with a leaning towards one language. The extent of this leaning is quite difficult to ascertain and is therefore an important potential source of error. Finally, the imperfection of the instrument per se contributes to error.

The concurrent validity coefficients for this study were obtained in the following way. First, the Greek and English scores for each subject were calculated and the Greek score was subtracted from the English score. Second, the subjects were classified as Greek dominant, balanced bilingual, or English dominant based on the difference calculated in the first step. A subject was classified

as Greek or English dominant if the difference between his or her Greek and English scores was larger than approximately ±1.5 standard errors of measurement, or ±5 points. If the score difference was smaller than the mentioned error of measurement difference, the subject was classified as a balanced bilingual. The selection of this value for the classification of the subjects was arrived at through a process of professional judgement and examination of the data.

The resulting classifications of subjects were then compared to the teacher ratings. The correlation coefficients obtained per grade and overall were as follows: Grade 1, .47; Grade 2, .47; Grade 3, .56; Grade 4, .27; and Grades 1-4, .47.

SUMMARY, CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS

Summary

The item pool of the instrument was developed during the spring of 1979, and the item selection process and evaluation was critiqued by over 30 bilingual doctoral students and language experts. Through this procedure the content validity of the instrument was established. The item pool was then pilot tested with 42 subjects in Grades 1 through 4 of the Greek bilingual schools in Tarpon Springs, Florida, at the end of May 1979. An item analysis of the data revealed the strong as well as the weak points of the instrument. In general, findings revealed (see Table 3) that the item pool's reliabilities were generally high, but that certain items, especially in the English part, were easy. These items were, therefore, substituted with others of better discriminating value.

The field-test administration that followed in early June included three schools (two in New York and one in Chicago) and an initial population of 361 subjects attending Grades 1-4. The item pool was administered to the subjects twice to examine both its test and test-retest reliabilities. Furthermore, teacher ratings of the subjects' language dominance were obtained. Such ratings were later compared with the test results in order to examine the

concurrent validity of the final instrument. The analysis of test reliabilities, conducted according to grade and school as well as across schools and grades (see Table 3), showed that reliabilities for all segments of the item pool (Subtest A, Subtest B, and Total Test) were quite high, both across grades and schools (.94 for the English and .95 for the Greek part).

After the internal consistency reliability of the item pool was established, selections of the best 40 items for the Greek and for the English were made. These items constituted the final instrument. Selection of these items was based primarily on the point biserial values. Items were selected that had large point biserial values on both the English and Greek versions of the test.

The final instrument (see Appendix A) was then examined for its internal consistency and other pertinent statistics. The results (see Table 4) showed that the internal consistency reliabilities, both by grade and overall, were high for both components of the instrument (.92 for the English vs. .93 for the Greek) and only slightly lower than those of the 80-item pool. They also showed that the reliabilities of the difference, which are crucial to the effectiveness of the instrument, were quite large and stable across grades. The instrument was then examined for its test-retest reliability to determine its stability over a period of time. This examination (see Table 5) showed that the instrument generally yielded high reliabilities. The values of test-retest reliabilities were, however, lower than the internal consistency reliabilities (.85 for the English and .78 for the Greek), due to additional sources of error that normally enter into a test-retest reliability.

The instrument was also examined by means of a t-test, to determine if any statistically significant difference was evidenced in the reliability values by school or by sex of the subjects. The statistical analyses (see Tables 6 and 7) showed that such differences did not exist.

Finally, the instrument was examined for its concurrent validity. Thus, the teacher ratings of the subjects were compared to the differences in test scores of the same subjects, and a correlation coefficient was obtained. The results of this analysis showed considerable support for the concurrent validity of the instrument.

Conclusions

Examinations of the instrument in relation to field-test data, internal consistency reliability, test-retest reliability, reliability by school and sex, and validity demonstrated that the instrument possessed high reliability and usable validity values. Furthermore, the data from the instrument was consistent with existing theories about the relationship between the language ability of subjects of different socioeconomic backgrounds, as well as the well established fact of difference in language aptitude between males and females.

These facts suggest that the instrument constitutes a respectable instrument for the measurement of the language dominance of Greek bilingual children attending Grades 1-4 in the United States.

Limitations of the Study and Recommendations

Despite careful consideration of possible sources of error, the study presents certain limitations. For example, the study does not provide for a parallel form which would be an additional measurement of the instrument's reliability and would have been of great assistance to school personnel and test administrators.

A more serious limitation, however, is the circularity appearing in the instrument characteristics. Ideally, selection of the best 40 items to build the final instrument should have been followed by a new administration of these items to the subjects to obtain a more accurate picture of the statistical characteristics of the instrument. Such an administration, however, was not feasible due to time constraints and problems of school approval for additional testing.

Another limitation of the instrument is also the assumption that the difficulty of its English part is equal to that of the Greek part. This assumption is primarily based on the fact that the items selected came from the same sampling domain and were judged to be of equal difficulty. Additionally, the administration of an instrument during the end of the school year does not meet ideal test conditions, since school personnel and subjects are more concerned with year-end tasks. A future researcher should try to administer such an instrument at a more propitious time of the school year. Future research should also include a more balanced subject population within grades, especially for the fourth grade, where 47 out of the 69 subjects come from one school only.

Finally, it is recommended that future research should attempt to expand the scope of the present instrument to include all elementary school grades, thus providing the school systems of the country with a reliable and valid instrument for the measurement of the language dominance of Greek bilingual children.

References

Anastasi, A. Psychological testing. (4th ed.). New York: MacMillan, 1976.

Andersson, T., & Boyer, M. Bilingual schooling in the United States. Washington, D.C.: U.S. Government Printing Office, 1970.

Aspira, Inc. v. Board of Education, No. 72 Civ. 4002 MEF (S.D. N.Y., September 20, 1974).

Bernardoni, Louis C. The testing of bicultural children. Sharing Ideas, 1967, 4 (2), 1-5.

Berry, G. L., & Lopez, C. A. Testing programs and the Spanish-speaking child: Assessment guidelines for school counselors. The School Counselor, 1977, 24, 261-269.

A Better chance to learn: Bilingual-bicultural education. Washington, D.C.: United States Commission on Civil Rights Clearinghouse (Publication 5), May 1975.

Bhaerman, R. Bilingual education: Its promises and problems. American Teacher, 1974, 62, 16-18.

Blanco G. et al. Bilingual education: Current perspectives. Arlington, Virginia: Center for Applied Linguistics, 1977. (ERIC Document Reproduction Service No. ED 146 822).

Burt, M. & Dulay, C. Some guidelines for the assessment of oral language proficiency and dominance. TESOL Quarterly, 12 (2), 177-193.

Cervenka, E. J. Administration manual for tests of basic language competence in English and Spanish, level I and II. Austin, Texas: Texas University, 1968. (ERIC Document Reproduction Service No. ED 027 063 & 027 064).

Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Coleman, J. et al. Equality of educational opportunity. Washington, D.C.: U.S. Government Printing Office, 1965.

Combs, A. W., Soper, D. W., & Courson, C. C. The measurement of self concept and self report. Educational and Psychological Measurement, 1963, 33, 493-500.

Cooper, R. Two contextualized measures of degree of bilingualism. Modern Language Journal, 1969, 53, 172-178.

Cooper, R. et al. Listening comprehension in a bilingual community. Modern Language Journal, 1969, 53, 235-241.

Diana v. California State Education Department, C.A. No. C-7037 RFD (N.D. Cal., February 3, 1970).

Edelman, M. The contextualization of school children's bilingualism. Modern Language Journal, 1969, 53, 179-182.

Eells, K. et al. Intelligence and cultural differences. Chicago: University of Chicago Press, 1951.

Ervin, S. Semantic shift in bilingualism. Journal of Psychology, 1961, 74, 230-241.

Evans, J. et al. The S/ELPS language performance screening: Technical report/final report. Austin, Texas: Southern Educational Development Laboratory, 1976. (ERIC Document Reproduction Service No. ED 142 040).

Fishman, J. A. Bilingualism in the barrio. Bloomington, Indiana: Indiana University Publication, 1971.

Gardner, R. C., & Lambert, W. E. Language aptitude, intelligence, and second-language achievement. Journal of Educational Psychology, 1965, 56, 191-199.

Harber, J. R. The bilingual child with learning problems (1976). (ERIC Document Reproduction Service No. ED 143 149).

Helmer, S. Demonstration of assessment of language dominance of Spanish-speaking bilingual children. Occasional Papers on Linguistics. No. 1. Carbondale. Illinois: Southern Illinois University. 1977. (ERIC Document Reproduction Service No. ED 144 417).

James, P. James language dominance test. Austin, Texas: Learning Concepts, 1974.

Kobrick, J. The compelling case for bilingual children. Saturday Review, 1972, 55, 54-58.

Kolers, P. Bilingualism and information processing. Scientific American, 1968, 218, 78-84.

Krug, M. The melting of the ethnics: Education of the immigrants. Bloomington, Indiana: Phi Delta Kappa Educational Foundation, 1976.

Kuhlman, N. A. Language dominance testing: Some questions. San Diego: San Diego State University, 1978. (ERIC Document Reproduction Service No. ED 154 614).

Lambert, W. et al. Linguistic manifestation of bilingualism. American Journal of Psychology, 1959, 72, 77-82.

Lambert, W., and Moore, N. Word association responses: Comparisons of American and French monolinguals with Canadian monolinguals and bilinguals. Journal of Personality and Social Psychology, 1966, 3, 713-720.

Locks, N. A., Pletcher, B. A. & Reynolds, D. F. Language assessment instruments for limited-English-speaking students: A needs analysis. Washington, D.C.: National Institute of Education, 1978.

Luft, M. et al. Development of a test instrument to determine language dominance of primary students: Test of language dominance (TOLD). Albuquerque, New Mexico: Southwest Research Associates, 1977. (ERIC Document Reproduction Service No. ED 141 393).

Macnamara, J. The bilingual's linguistic performance: A psychological overview. Journal of Social Issues, 1967, 23, 58-77.

Moskos, C. C. Greek Americans: Struggle and success. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1980.

Oakland, T. Psychological and educational assessment of minority children. New York: Brunner/Mazel, 1977.

Politzer, P. L., and McKay, M. A pilot study concerning the development of a Spanish/English oral proficiency test. Stanford University, California: Stanford Center for Research and Development in Teaching, 1974. (ERIC Document Reproduction Service No. ED 108 460).

Purnell, R. F. Portuguese bilingual test development. Final report. Rhode Island: Rhode Island University, Kingston Curriculum Research and Development Center, 1973. (ERIC Document Reproduction Service No. ED 101 581).

Restak, R. M. The brain: The last frontier, New York: Doubleday & Company, Inc., 1979.

Rodriguez, C. The structure of failure. Urban Review, 1974, 7, 215-226.

Saville, M., and Troike, R. A handbook of bilingual education. Washington, D.C.: Teachers of English to Speakers of Other Languages, 1971.

Shitala, M. P., and Maure, H. Jr. The use of metropolitan readiness tests with Mexican-American children. California Journal of Educational Research, 1970, 21, 182-187.

Shore, M. Final Report of Project BEST: The content analysis of 125 Title VII bilingual programs funded in 1969 and 1970. New York City: Bilingual Education Applied Research Unit. Hunter College, 1974.

Shutt, D. Shutt primary language indicator. Phoenix, Arizona: Citizen's Press, 1974.

Skoczylas, R. Home bilingual usage estimate. Gilroy, California: Copywritten instrument, 1971.

Spolsky, B. et al. Three functional tests of language proficiency. TESOL Quarterly, 1972, 6, 221-235.

U.S. Bureau of Census. Washington, D.C.: U.S. Government Printing Office, 1976.

Vlachos, E. C. The assimilation of Greeks in the United States. Athens, Greece: National Center for Social Research, 1968.

William, C. The construction of standardized tests for Welsh-speaking children. Educational Research, 1971, 14, 29-31.

Woodford, P. The importance of testing. Arlington, Virginia: Center for Applied Linguistics, 1977. (ERIC Document Reproduction Service No. ED 146 822).

Zirkel, P. A. A method for determining and depicting language dominance. TESOL Quarterly, 1974, 8, 7-16(a).

Zirkel, P. A. Modular sequence: Teaching reading to bilingual learners. West Hartford, Connecticut: Hartford University, 1974. (ERIC Document Reproduction Service No. ED 106 239(b).

APPENDIX A

LANGUAGE DOMINANCE TEST FOR GREEK BILINGUAL CHILDREN: ENGLISH AND GREEK PARTS (FIELD TEST)

(Greek version omitted because of poor reproducibility)

Name_____ Date_____

Grade_____ Time Started_____

School_____ Time Completed_____

Directions to the Student:

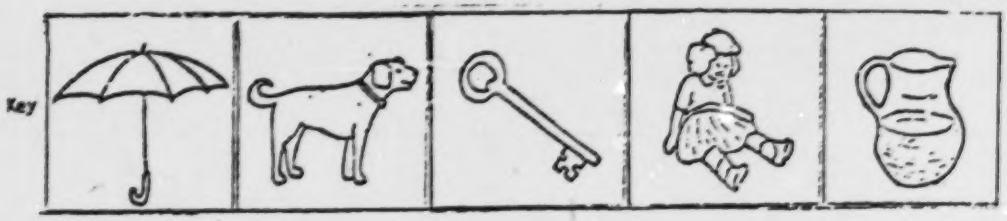
Read the directions to yourself as your proctor reads them out loud to you.

In the following exercise you will see five pictures on the right and a word on the left of each row. Put an X on the picture that means the same as the word on the left. BE SURE TO ANSWER ALL THE QUESTIONS.

EXAMPLE : I

You will see the word Bird on the left and an X on the picture of the bird. Now do the next example.

EXAMPLE: II



1. Cup

2. Brush

★ 3. Blouse

★ 4. Trousers

★ 5. Pail

★ Items included in the Final Instrument

6. Knife

7. Purse

8. Bat

9. Desk

10. Envelope

★ 11. Grapes

★ 12. Camera

★ 13. Nail

14. Pin

★ 15. Ham

★ 16. Bench

★ 17. Pick

18. Necklace

★ 19. Veranda

★ 20. Cone

21. Toes

★22. Rubbers

★23. Funnel

24. Casserole

25. Beans

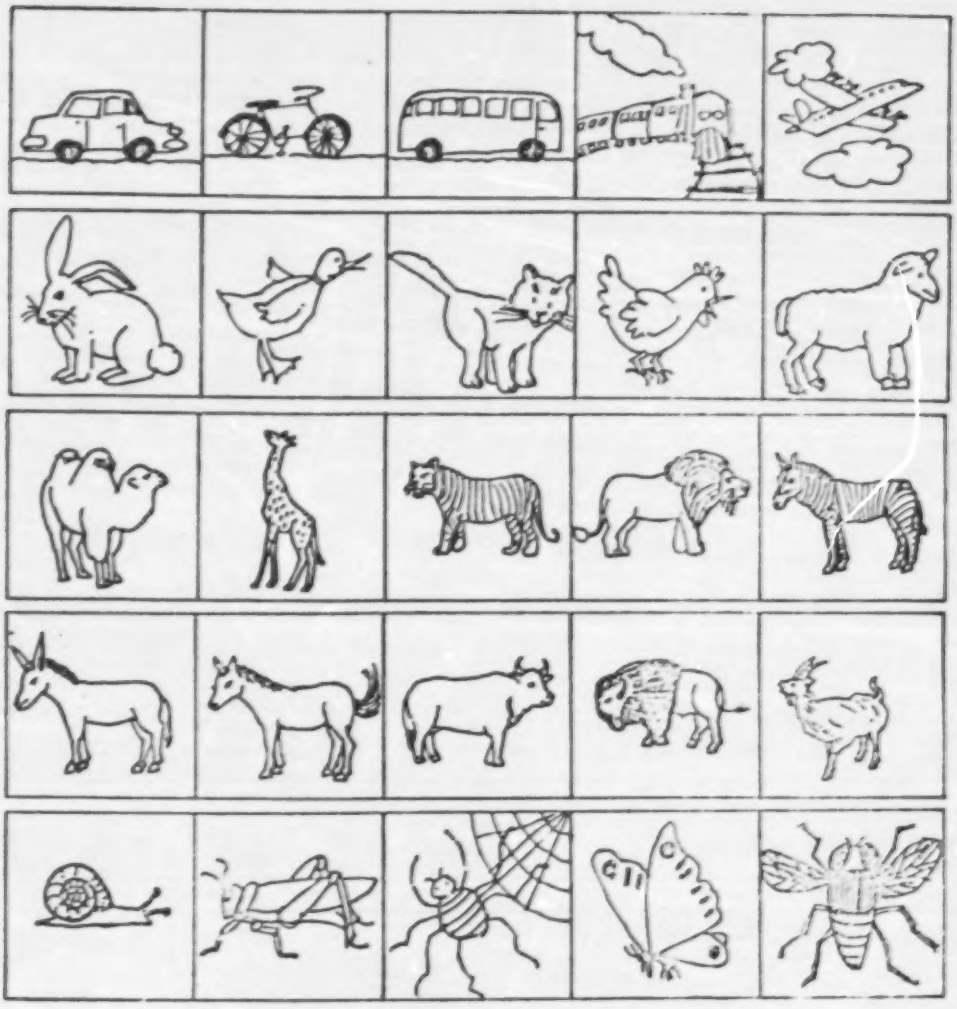
26. Train

★27. Hare

28. Zebra

29. Buffalo

30. Butterfly



31. Log

★ 32. Plumber

★ 33. Soldier

★ 34. Engineer

★ 35. Baker

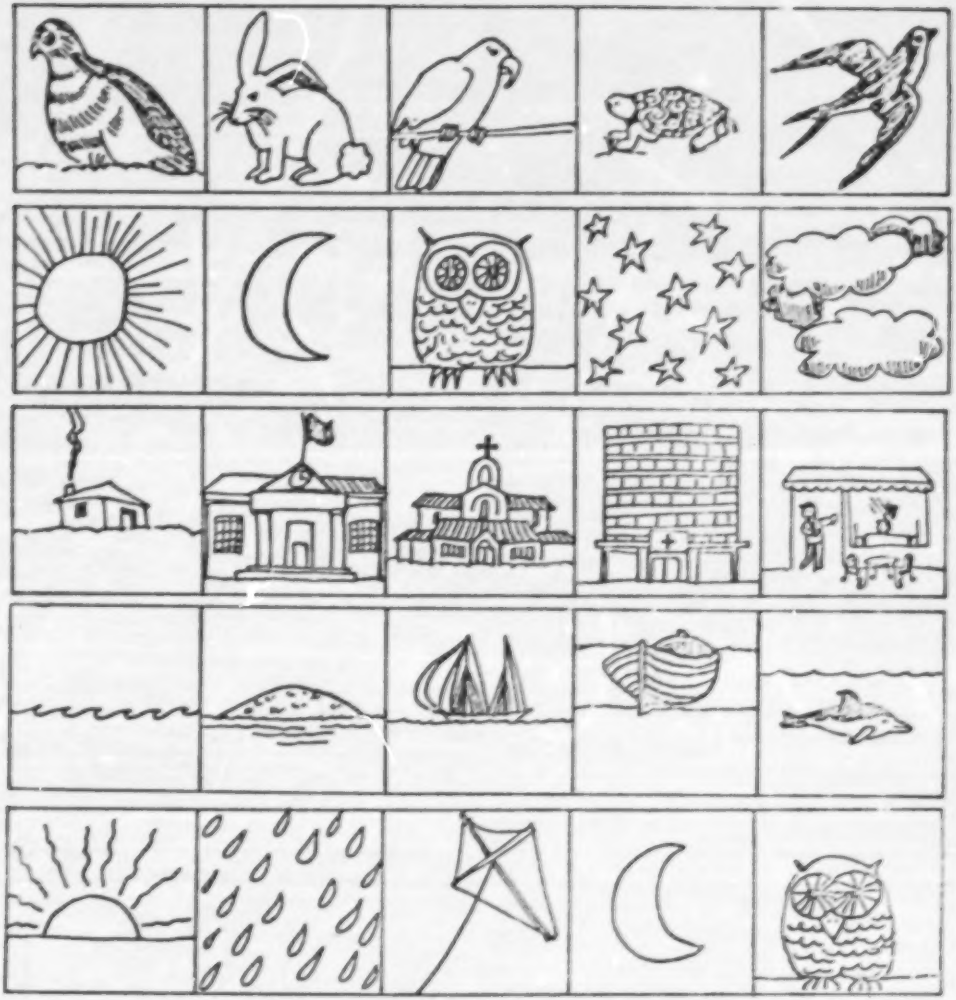
★ 36. Swallow

37. Rays

★ 38. Courthouse

★ 39. Shore

★ 40. Drops



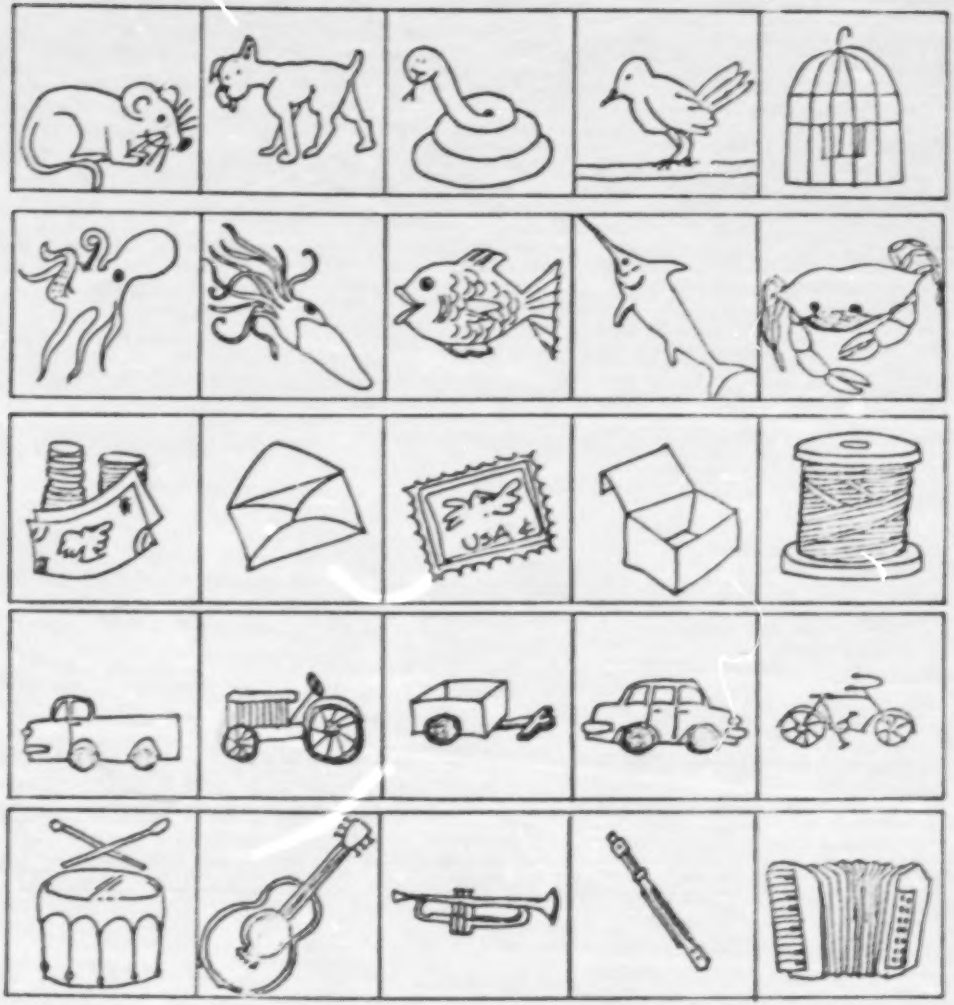
41. Snake

★42. Squid

★43. Carton

★44. Tractor

★45. Accordion



46. Slingshot

47. Stand

48. Spiral

★ 49. Crayons

★ 50. Painting

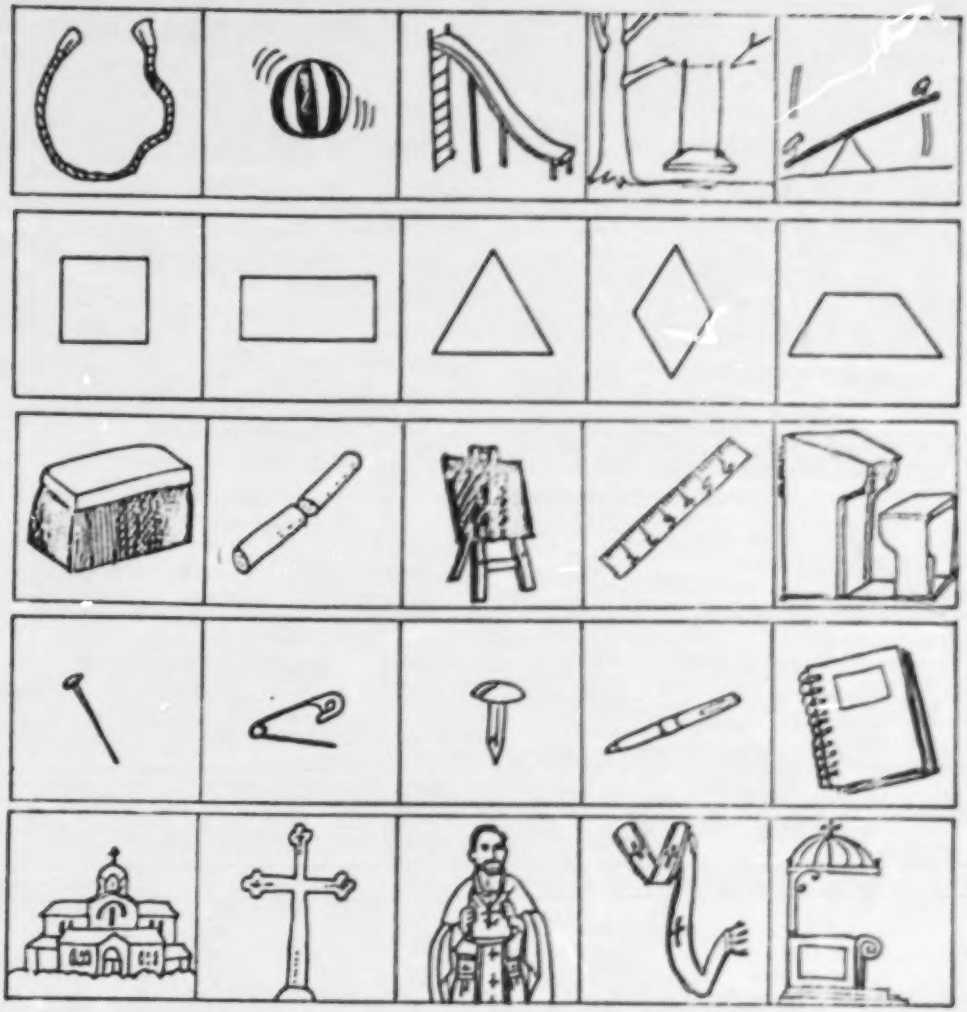
★ 51. Slide

52. Rectangle

★ 53. Ruler

54. Safety pin

55. Cathedral



56. Miter

57. Dome

58. Chandelier

59. Altar

60. Staff

Name:_____ Date_____

Grade:_____

School:_____

Time Started:_____

Time Ended:_____

DIRECTIONS

Please answer as many questions as you can in the space provided after each question. Do Not use a "YES" or "NO" answer. Your answers will not be seen or graded by your teacher.

EXAMPLE

QUESTION: What is your name?

ANSWER: My name is John or my name is Mary, etc.

QUESTIONS

1. Q. How old are you?
   A. _____
2. Q. Were you born in the United States?
   A. _____
★ 3. Q. What do you like about watching TV?
   A. _____
4. Q. Who is your favorite character in a comic book and why?
   A. _____
★ 5. Q. What sport do you like best?
   A. _____
★ 6. Q. Why would a dog make the best pet?
   A. _____
★ 7. Q. Which subjects do you like the most in school?
   A. _____
★ 8. Q. What would you like to do during the summer?
   A. _____
★ 9. Q. Where would you like to go for a field trip?
   A. _____

★ 10. Q. What would you expect to see in a museum?
A. _____

★ 11. Q. How do you celebrate your birthday?
A. _____

12. Q. If you had a lot of money, what would you do with it?
A. _____

★ 13. Q. Which occupation do you like the most?
A. _____

14. Q. How would you like to utilize your free time?
A. _____

15. Q. What is a bird of prey?
A. _____

16. Q. What form of transportation do you use to come to school every day?
A. _____

17. Q. What are the advantages of living in a big city instead of a small town?
A. _____

18. Q. How would you like to spend your time, if you didn't have to go to school every day?
A. _____

19. Q. If you could change into an animal, which one would you like to be and why?
A. _____

★ 20. Q. If you could live in another country, which one would you choose and why?
A. _____